

The Russian Playbook: An Assessment of EUvsDisinfo Data

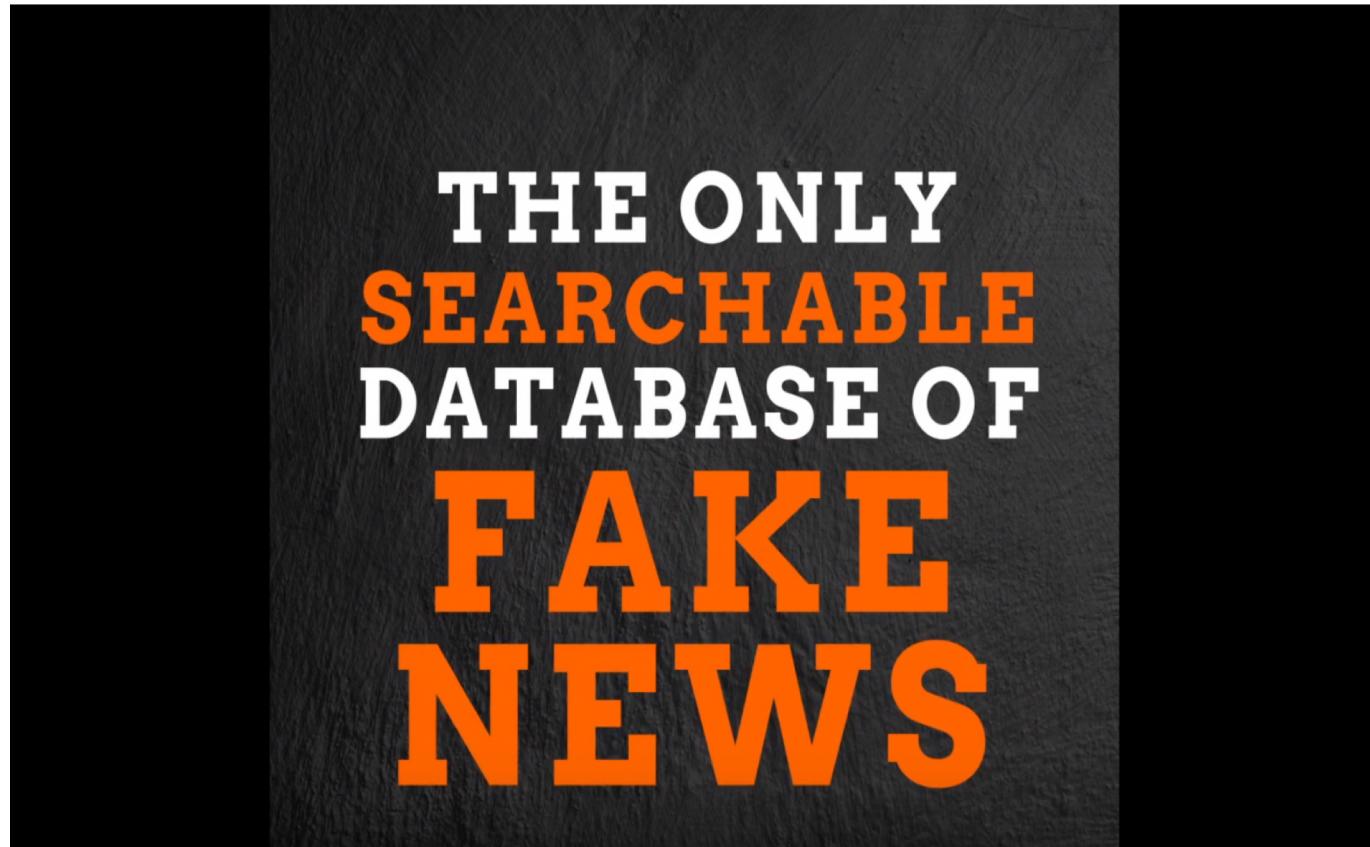
By Lauren Barden-Hair

University of California
Berkeley

May 6, 2020



What is this dataset?



- 1) "EUvsDisinfo is the flagship project of the European External Action Service's [East StratCom Task Force](#)"
- 2) The database is updated weekly with identified and debunked disinformation cases.
- 3) EUvsDisinfo was established in 2015. It initially began by looking only at European countries, but has since expanded its scope. The database currently has over **8390 cases**.

What does a disinfo “claim” look like?

Claim

DISINFO: COOPERATION BETWEEN THE MEDITERRANEAN COUNTRIES THREATENS GERMANY'S HEGEMONY

SUMMARY

Summary of claim content

The cooperation between the Mediterranean countries might become a problem for securing Germany's hegemony over the [European] continent. Of course an issue is that the interests of Spain, Portugal and Greece not always coincide, but all these countries have suffered greatly from the fiscal regime imposed by Germany and its financial policies.

The pandemic has demonstrated that Germany, in a time of crisis, does not wish to help its partners in the European Union and is first and foremost preoccupied with its own interests. This has been the same all the time, but in more peaceful years, the illusion of European unity has been stronger.

DISPROOF

Debunking of the narrative

This claim exploits the coronavirus pandemic to push a recurring pro-Kremlin disinformation narrative about "German hegemony in Europe" and the EU's alleged imminent collapse.

This narrative also aims to pit EU member states against each other in an attempt to undermine European solidarity, another recurring pro-Kremlin disinformation tactic.

PUBLICATION/MEDIA

→ News Front - Russian (Archived)

REPORTED IN:
Issue 194

DATE OF PUBLICATION:
20/04/2020

LANGUAGE/TARGET AUDIENCE:
Russian

COUNTRY:
EU, Portugal, Greece, Germany, Spain

KEYWORDS:
coronavirus, EU, EU disintegration, Mediterranean

Name of Publication/Outlet

Day/Month/Year

Language of article

Countries mentioned

Keywords

Motivating Puzzles

- 1) Who is Russia targeting the most with these disinformation narratives?
- 2) What issues are most commonly raised in disinformation articles?
- 3) Are there more disinformation articles surrounding national elections?
- 4) Has there been a change in quantity of disinformation articles over time?

Overview of My Comp Tools Project

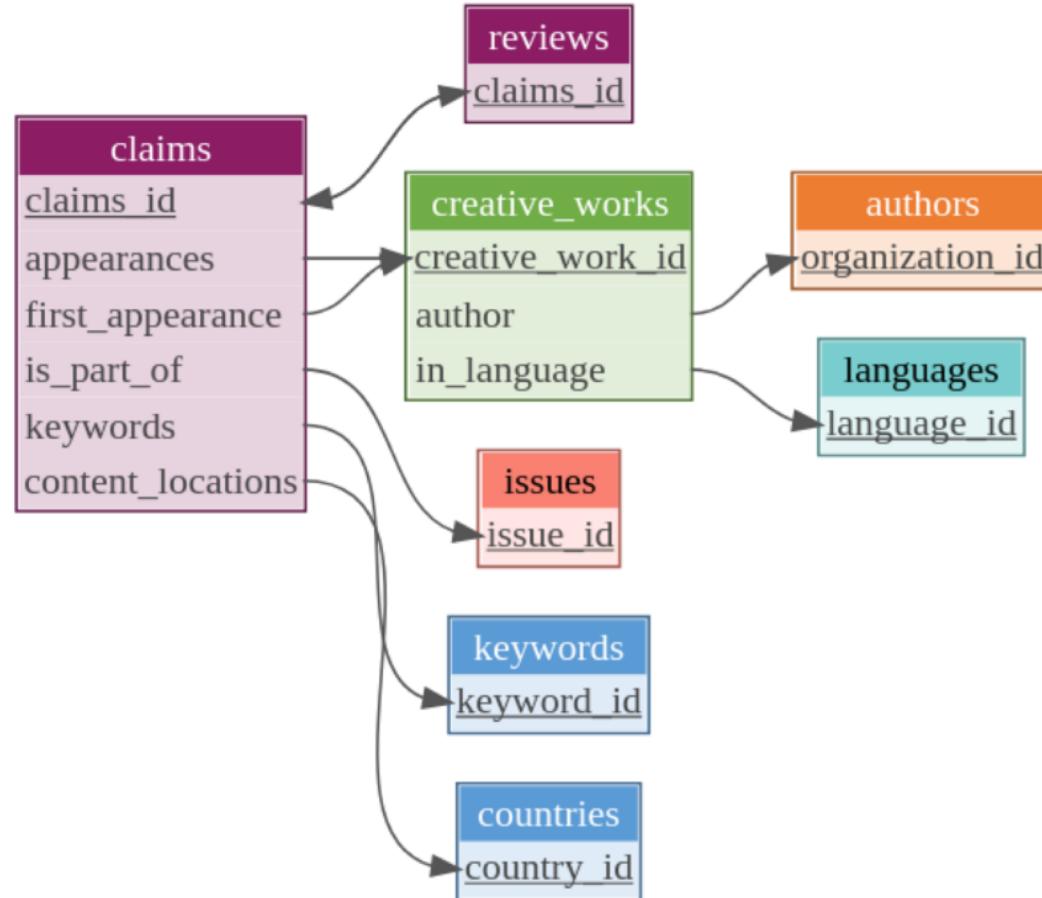
- How I gathered data
- How I cleaned/manipulated the data for basic analysis
- Preparing the data for visualizations
- Visualizing the data
- Findings, challenges of the project, and next steps



Main Tools used in this Project:

For this project, I worked in R and utilized EUvsDisinfo's API. I also made use of dplyr, ggplot, tidyverse, and other packages.

Gathering Data: Working with the API



Manipulating/Analyzing the Data

Example observation. Note that some variables are lists (like "Keywords"):

Claim	Claim ID	Date	Keywords	Languages
EU wants control of Crimea	--	05/04/2020	Ukraine, Crimea, NATO, Zelensky, Russophobia	Russian, Estonian, Latvian

Counting by Keyword, Location, and Language:

Ukrainian	Crimea	NATO	Zelensky	Russophobia
897	420	555	79	115

Preparing Data for Analysis/Visualization

Generated “Regions” Category
for Languages:

- Middle East
- Eastern Europe
- Western Europe
- US and UK
- Central Asia

Generated Common “Topics”
Category for Keywords:

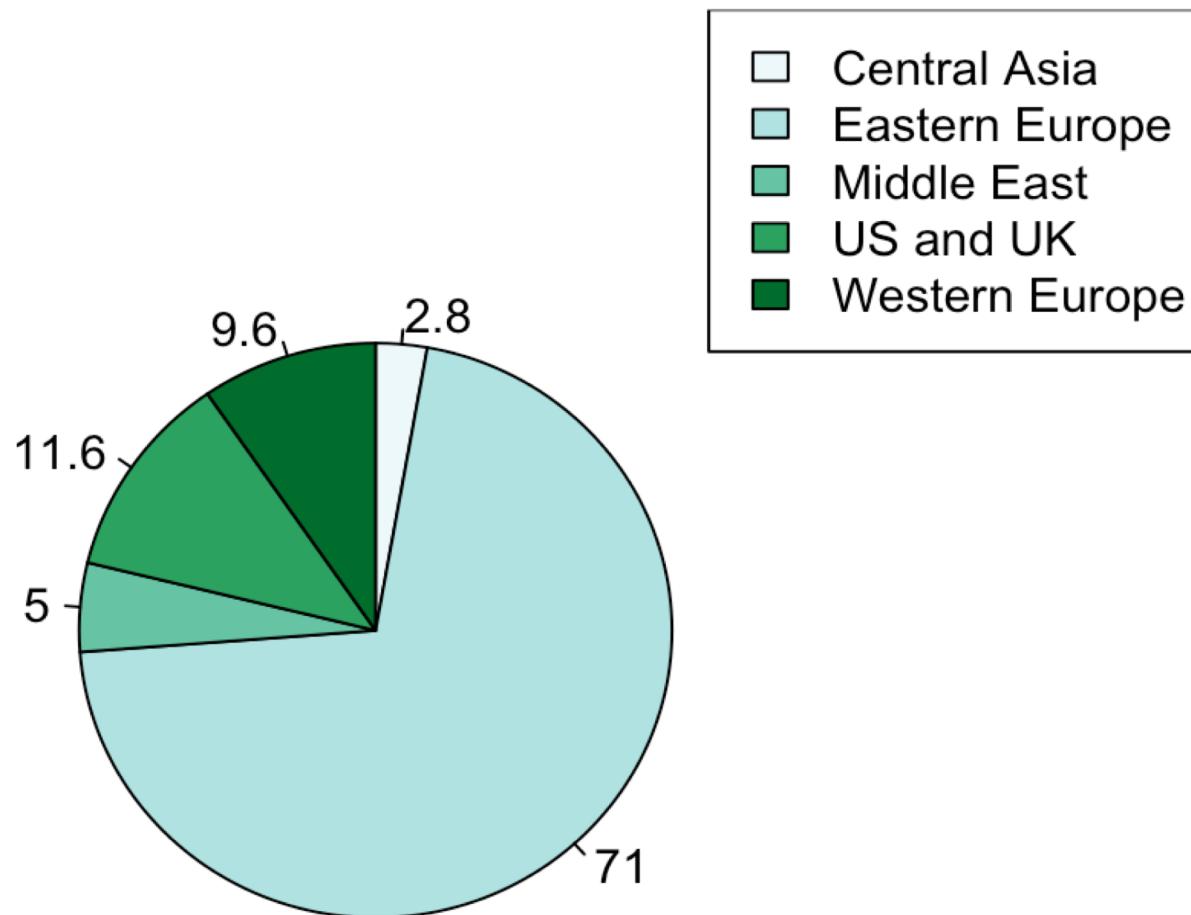
- Eastern Europe
- Ukraine
- EU/NATO
- Elections
- US
- Russia
- Syria/Middle East

Grouped Items by Year to
see the Effect Over Time:

- 2015
- 2016
- 2017
- 2018
- 2019
- 2020

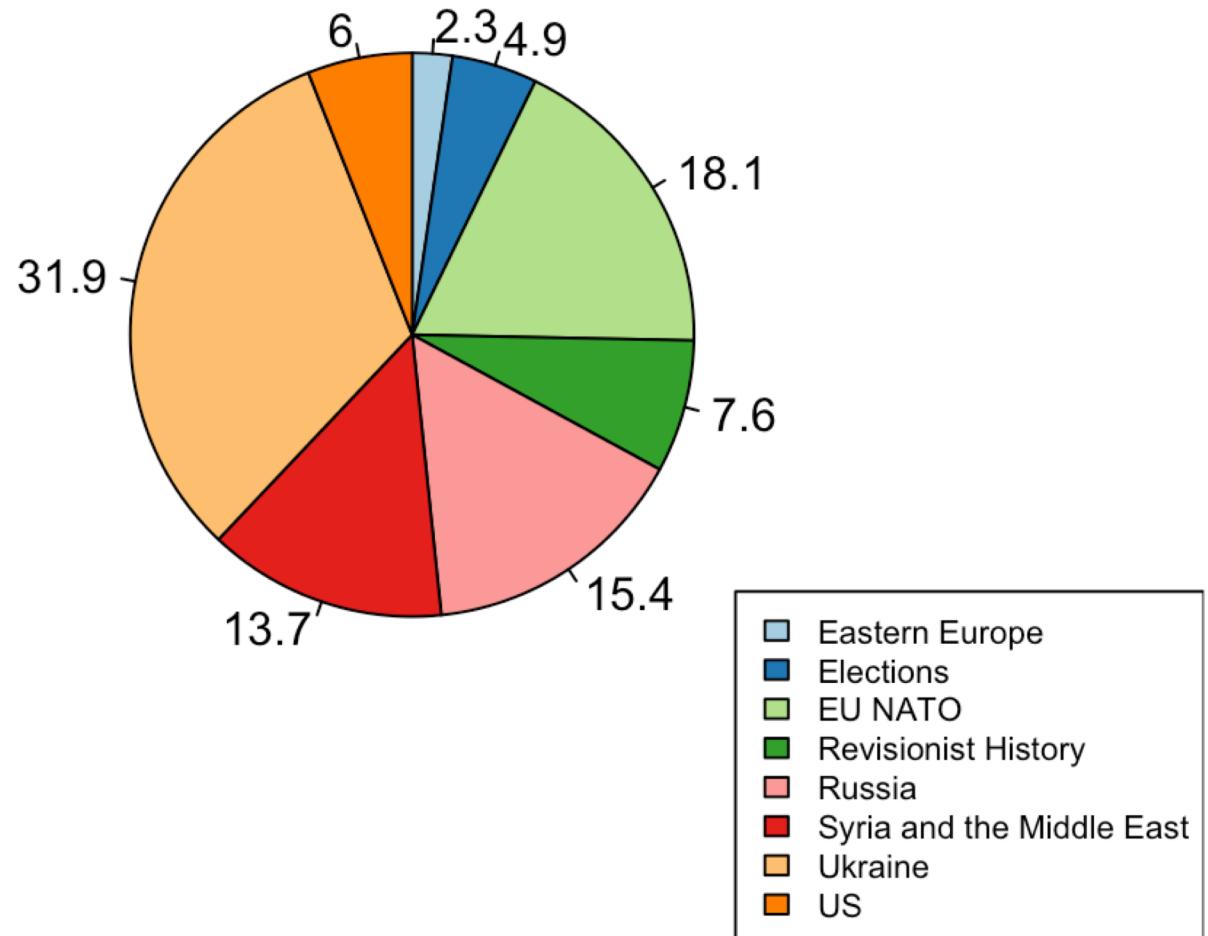
What Regions are Targeted the most by Russian Disinformation?

Pie Chart of Region Frequencies by Percent



Pie Chart of Topic Frequencies by Percent

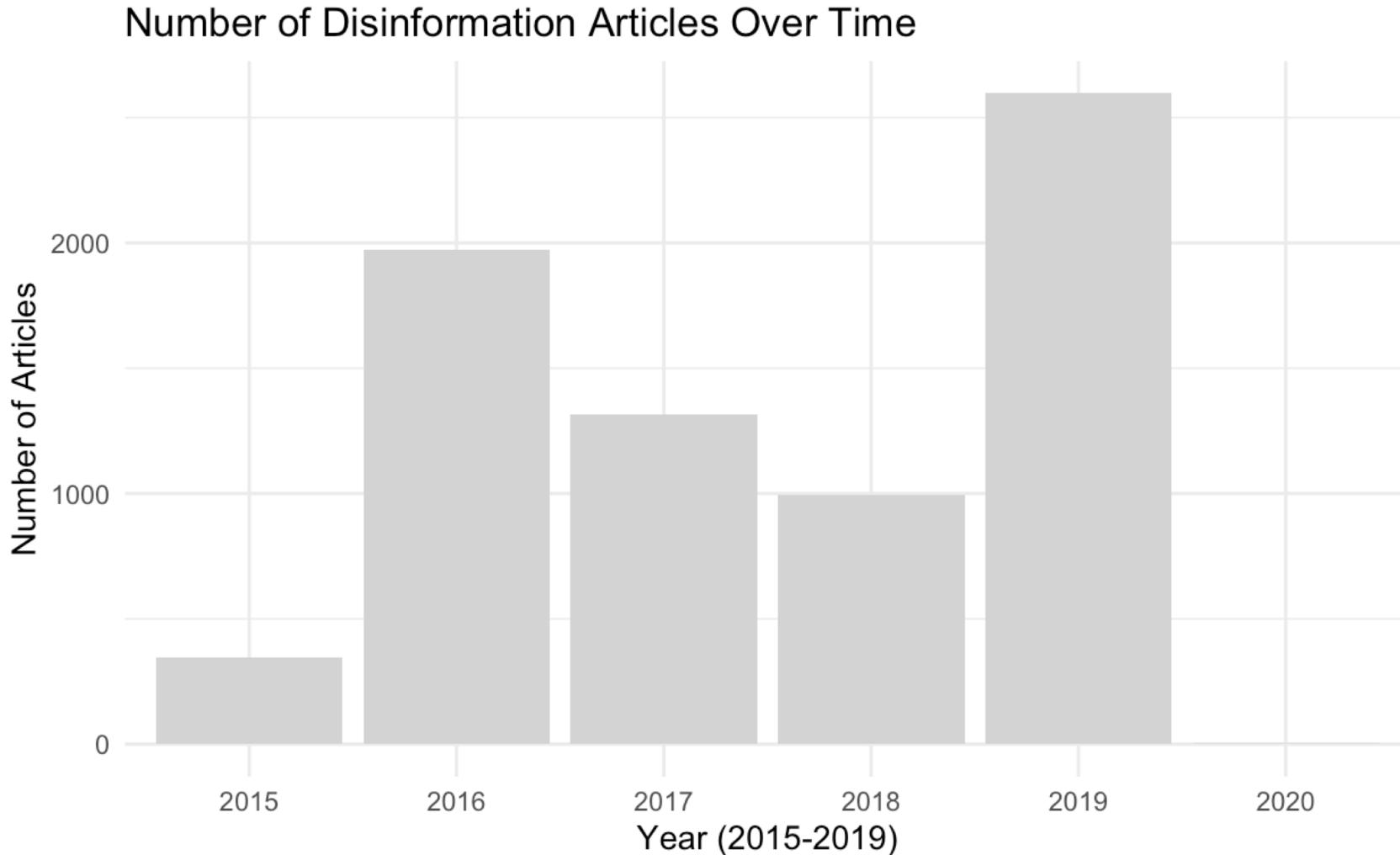
What are the most common topics addressed in Russian disinformation?



Is there an increase in disinformation frequency in an election year?



Are Russian disinformation attempts increasing over time?



What was most challenging about the project?

The complexity of the API data format was the biggest challenge of the project

I solved this problem by flattening the data and utilizing the lapply function. I also used the hash /dictionary package to tally the string variables



Next Steps

- 1) Write a new codebook to account for both tone and different content categories (not just keywords, but more IR categories [ex: economic, security, nuclear, soft power, etc.])
- 2) Need to look at websites distributing the articles as a better proxy for target audience
- 3) More specificity on frequency surrounding election months or weeks, instead of simply looking at years

